Verlag Schnelle
Eberhard und Wolfgang Schnelle GmbH & Co
2085 Quickborn bei Hamburg
Postscheckkonto Hmb 103676. Global Bank Hamburg 25011
Druck: S. Maurischat & A. Bevensee, Quickborn. Printed in Germany.

GRUNDLAGENSTUDIEN

AUS

KYBERNETIK

UND GEISTESWISSENSCHAFT

BAND 6 HEFT 1 MÄRZ 1965

Kurztitel: GrKG 5(3/4) Schnelle, 2085 Quickborn/Germany

INHALT

VERLAG SCHNELLE, QUICKBORN BEI HAMBURG

Neuerdings vollzieht sich eine immer stärker werdende Annäherung zwischen Natur- und Geisteswissenschaft als Auswirkung methodologischer Bestrebungen, für die sich das Wort Kybernetik eingebürgert hat. Die Einführung statistischer und speziell informationstheoretischer Begriffe in die Ästhetik, die invariantentheoretische Behandlung des Gestaltbegriffs und die Tendenzen, zwischen der Informationsverarbeitung in Maschine und Nervensystem Isomorphismen nachzuweisen, sind nur drei Symptome dafür.
Die Grundlagenstudien aus Kybernetik und Geisteswissenschaft sollen der raschen Publikation neuer Resultate dienen, welche diese Entwicklung zu fördern geeignet sind. Veröffentlicht werden vor allem grundlegende Ergebnisse, sowohl mathematischer, psychologischer, physiologischer und in Einzelfällen physikalischer als auch philosophischer und geisteswissenschaftlicher Art. Nur in Ausnahmefällen werden dagegen Beiträge über komplexere Fragen der Nachrichtentechnik, über Schaltungen von sehr spezieller Bedeutung, über Kunst und literaturgeschichtliche Probleme etc. angenommen. In geringer Zahl werden Buchbesprechungen veröffentlicht. (GrKG 1, 1960, S. 1)

Erscheinungsweise: Viermal im Jahr mit je 32 bis 48 Seiten.
Beiheft: Im Jahr erscheint für Abonnenten ein Beiheft.
Preis: DM 4,80 je Heft und Beiheft. Für Angehörige von Lehranstalten 2,88 DM.
Im Abonnement Zustellung und Jahreseinbanddeckel kostenlos. Bezug durch Buchhandel oder Verlag.
Manuskriptsendungen: an Schriftleitung gemäß unserer Richtlinien auf der dritten Umschlagseite.

Schriftleitung

Prof. Dr. Helmar Frank
Institut für Kybernetik
Berlin 46, Malteserstr. 74/100

Les sciences naturelles et les sciences humaines se rapprochent de plus en plus; ce rapprochement est une conséquence des tendances métodologiques appelées cybernetique. L'introduction en esthétique de termes statistiques et surtout de termes de la théorie de l'information, le fait de considérer mathématiquement la notion de Gestalt comme une invariante, et les tendances à chercher des isomorphismes entre la transformation de l'information par les machines et par le système nerveux sont seulement trois exemples du dit rapprochement.
Les «Grundlagenstudien aus Kybernetik und Geisteswissenschaft» ont pour but de publier rapidement des résultats nouveaux capables de contribuer à ce dévéloppement. Surtout des résultats fondamentaux (soit de caractère mathématique, psychologique, physiologique et quelquefois physique — soit de caractère philosophique ou appartenant aux sciences humaines) sont publiés. Par contre des travaux concernant soit des questions assez complexes de la théorie de communication et télécommunication, soit des reseaux éléctriques ayant des buts trop spéciaux, soit des problèmes de l'histoire de l'art et de la litérature etc. ne sont acceptés qu'exceptionnellement aussi que les comptes rendus de nouveaux livres. (GrKG, T. 1, 1960, p. 1.)

Il paraissent 4 numéros de 32 à 48 pages par an et un numéro spécial, pour les abonnés. Prix: DM 4.80 le numéro (et le numéro spezial); pour membres des universités et écoles DM 2.88. L'envoi et la couverture du tome complèt (à la fin de chaque année) est gratis pour les abonnés.
Les GrKG sont vendus en librairie ou envoyés par les Editeurs Schnelle
Les manuscrits doivent être envoyés au rédacteur en chef. Quant à la forme voir les remarques à la page 3 de cette couverture.

Rédacteur en chef

Prof. Dr. Helmar Frank
Institut für Kybernetik
Berlin 46, Malteserstr. 74/100

Natural and cultural sciences are in train to come together closer and closer as a consequence of methodologica-tendencies called cybernetics. The introduction of terms of statistics and specially of information theory into the terminology of esthetics, the interpretation of 'Gestalten' as mathematical invariants, and the search for isomorphisms by comparing information handling in computers and the brain are only three symptoms of the process mentioned above.
The Grundlagenstudien aus Kybernetik und Geisteswissenschaft would like to cultivate this tendencies by rapid publication of new results related to cybernetics, especially results of basic interest, no matter whether belonging to the field of mathematics, psychology, physiology and sometimes even of physics, or rather to the fields of philosophy and cultural sciences. But papers which concern complex technical problems of transmission and processing of information, or electrical networks with very limited purpose, or the history of art and literature, are accepted only exceptionally. There will also be few recensions of books. (GrKG, 1, 1960, p. 1)

GrKG are published in 4 numbers each year, with 32–48 pages per number. A special number is edited each year for the subscribers.
Price: DM 4.80 per number (and spezical number). For members of universities and schools DM 2.88. Mailing and cover of the volume (to be delivered together with the last number each year) is free for subscribers. The GrKG may be received by booksellers or directly by the publisher.
Papers should be sent to the editors. For the form of manuscript see page 3 of this cover.

Editor

Prof. Dr. Helmar Frank
Institut für Kybernetik
Berlin 46, Malteserstr. 74/100

ÜBER DIE VERTEILUNG DER SEMANTISCHEN INFORMATION IN LEHRPROGRAMMTEXTEN

von Herbert Anschütz, Heidelberg

1. Das (m,i)-Diagramm

Wenn man die Lehrbegriffe eines Lehrprogrammtextes in der Reihenfolge ihres Auftretens numeriert und diese Begriffsnummern m über den Framenummern i aufträgt, so erhält man ein Diagramm, welches den Begriffsfortschritt des Lernens mit dem Framefortschritt zeigt. Bei den meisten untersuchten Programmen ergab sich, daß der mittlere Begriffsfortschritt/Frame eine Konstante v über das ganze Programm hin war, d.h. die Anzahl m der gebrachten Begriffe ist eine recht genaue lineare Funktion der Framezahl i:

$$m = z' + vi \quad .$$

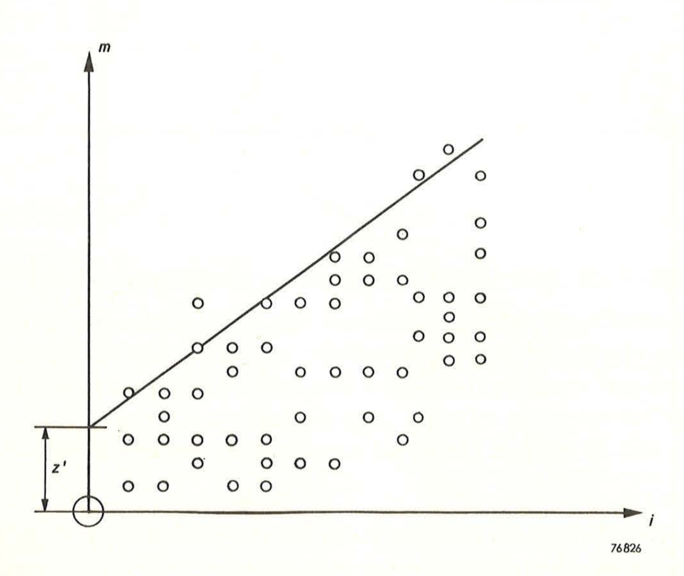


Bild 1: (m,i)-Diagramm

Der Autor verwendet hier zur Abkürzung die aus dem angelsächsischen Schrifttum übernommene Terminologie, d. h. "Frame" statt "Inhaltseinheit" bzw. "Lehrschritt".

Dabei ist z' eine Konstante, die die Dimension einer Anzahl von Begriffen/Frame hat, v ist der mittlere Begriffsfortschritt/Frame.

Wenn man nun nicht annehmen will, daß die Begriffszahl/Frame immer größer wird, (d. h. die Frames immer länger), so muß es eine zweite Begrenzung des markierten Feldes, d. h. ein m_h geben, für das für $i > a$ der Begriff $m < m_h$ nicht mehr genannt wird (weil er dann eingeprägt ist oder ein neuer Oberbegriff erarbeitet wurde). Das markierte Feld liegt zwischen zwei parallelen Geraden $m = z' + vi$ und $m_h = z'' + vi$. Dabei ist z'' ebenfalls eine Begriffszahl. $z = z' + z''$ ist die Anzahl der Begriffe, die zwischen der Vorderkante m und der Hinterkante m_h des Erwähnungsfeldes liegen.

$$m - m_h = z' + z'' = z.$$

Ein Punkt (m, i) stellt eine Erwähnung des m-ten Begriffes im i-ten Frame dar. Die beschriebene Darstellung des Programminhaltes nennen wir das (m, i)-Diagramm des Lehrprogrammes.

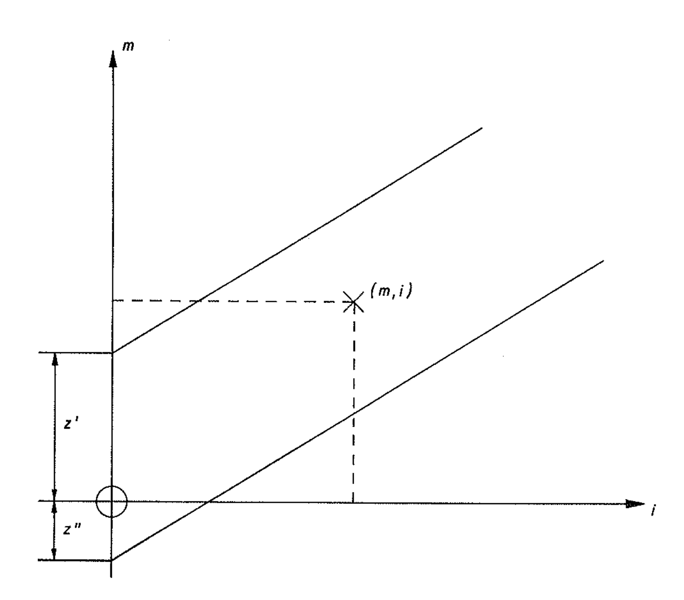


Bild 2 Erwähnung, Begriffsvorgriff und Begriffsrückgriff im (m, i)-Diagramm

2. Die relative mittlere Begriffsredundanz

Alle Erwähnungen eines Begriffs m liegen auf einer Parallele zur i-Achse und stellen Repetitionen des Begriffes dar. Vom Frame i bis zum Frame k wird der Begriff immer wieder repetiert, d. h. redundant erwähnt. Dabei werden Urteile zu den Begriffen gebildet, die in den Feldern A und B erwähnt sind. Wenn

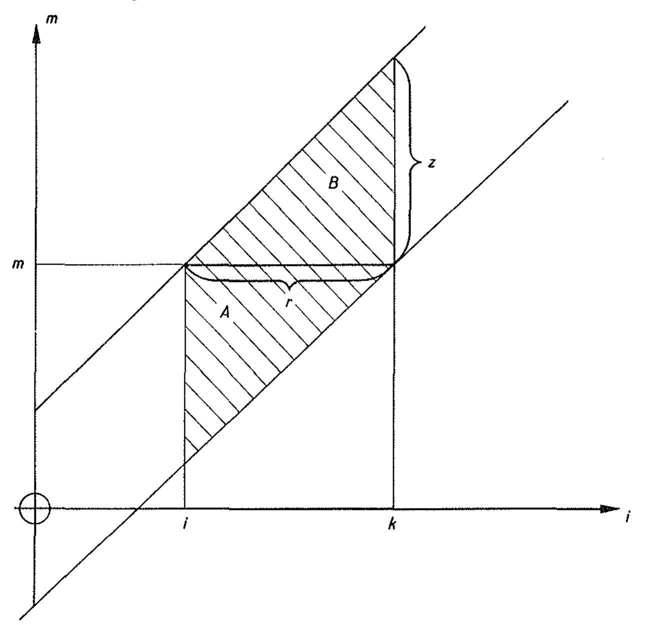


Bild 3 Begriffszahl z und Repetitionszahl r im (m, i)-Diagramm

ein Frame präsentiert wird, wird jeweils der gesamte Begriffsinhalt einer senkrechten Schnittlinie durch das Markierungsfeld ins Bewußtsein aufgenommen. Das sind jeweils z Begriffe. Von diesen Begriffen werden v zum ersten Male erwähnt, z - v werden redundant erwähnt, sind also schon einmal vorgekommen. Die relative Redundanz der im Frame erwähnten Begriffe ist also

$$\varrho = \frac{z - v}{z} \quad .$$

Dabei hebt sich der mittlere Informationsgehalt der Begriffe fort, so daß diese Zahl die gleiche ist, die man auch durch Ermittlung des Informationsgehaltes in bit errechnen könnte. $\beta = v/z$ ist die relative Negentropie, d. h. der mittlere bezogene Informationsgehalt des Frames. Es gilt wie üblich: $\beta + \varrho = 1$. β könnte auch als der Neuigkeitswert des Frames bezeichnet werden. Die beiden Flächenanteile A und B von Fig. 3. haben zusammen die Größe $A + B = rz$. r ist die Zahl der Frames, in denen ein Begriff erwähnt wird, d. h. also $v = z/r$. z ist die Zahl der Begriffserwähnungen in einem Frame, d. h. die Zahl der verschiedenen erwähnten Begriffe. (Wenn ein Begriff im Frame mehrmals erwähnt ist, so gerät es doch im semantischen Zusammenhang nur einmal ins Bewußtsein. Mehrfache

Erwähnung einer Begriffsbenennung im Frame oder Erwähnungen durch Relativ- oder Reflexivpronomina sind syntaktische Mittel der Sprache, wenn ein Frame regelrecht strukturiert ist, d. h. eine Sinneinheit bildet. Eine Begriffserwähnung wird deswegen in einem Frame nur einmal gezählt.) Die Dimension von z ist also

$$[z] = \text{Begriffserwähnungen/Frame} \quad .$$

Die Dimension von rz ist damit

$$[rz] = \text{Begriffserwähnungen} \quad .$$

Ein Begriff wird mit rz anderen Begriffen zusammengebracht und selbst r mal wiederholt. Dadurch werden für jeden Begriff eine Anzahl $u < rz$ logisch definitorische Urteile mitgeteilt. Aus $\beta = v/z$ und $v = z/r$ folgt $r = 1/\beta$, d. h. es gilt der Satz:

> Die mittlere reziproke relative Negentropie eines Frames ist gleich der mittleren Anzahl der Repetitionen eines Begriffs.

3. Bemessung von Programmlängen

Wenn ein Lehrstoff, der aus M Begriffen besteht, in I Frames gelehrt wird, so ergibt sich bei gleichmäßigem Fortschreiten ein Begriffsfortschritt:

$$v = \frac{M}{T} \text{ Begriffe/Frame} \quad .$$

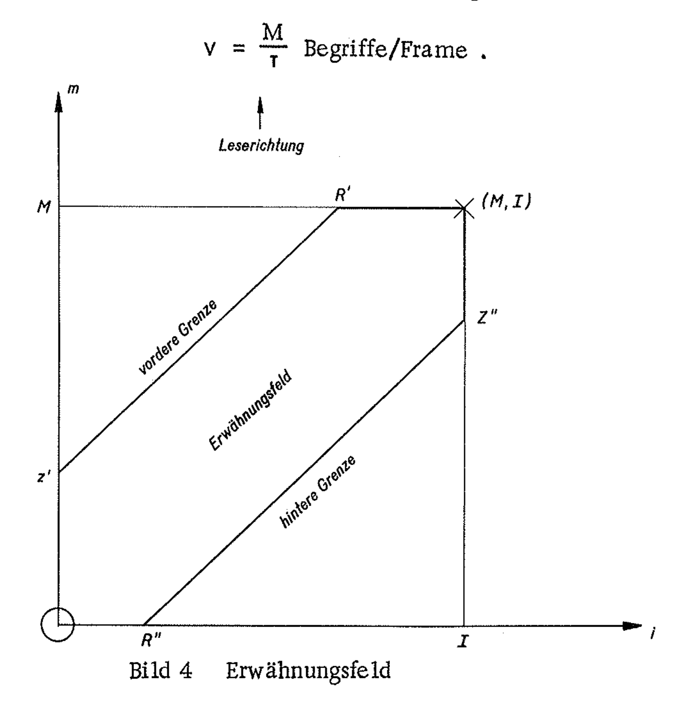


Bild 4 Erwähnungsfeld

Die vordere Grenze des Erwähnungsfeldes ist damit gegeben durch

$$m = \frac{M}{I} i + z'.$$

z' nennen wir den mittleren Begriffsvorgriff. Die hintere Grenze des Erwähnungsfeldes ist

$$m = \frac{M}{I} i - z'' \quad \text{mit dem Begriffsrückgriff } z''.$$

Alle Erwähnungen der Begriffe des Lehrstoffes (Lehrbegriffe) liegen in dem Sechseck (O, R", Z", P, R', z').

Wegen $v = z/r$ ist die Anzahl I der für M benötigten Frames

$$I = M/v = M\, r/z$$

Wenn man in diese Gleichung

$$r = \frac{1}{1 - \varrho}$$

einführt, erhält man

$$I = \frac{M}{z(1 - \varrho)} .$$

Das ist die Anzahl der Frames einer durch z und ϱ gekennzeichneten Art, mit der ein Lehrstoff M behandelt werden muß. Nun darf die Begriffszahl z eines Frames einen gewissen Wert z_o nicht überschreiten, weil die Kapazität des Gegenwartsspeichers (Bewußtsein) mit $K_k \approx 160$ bit beschränkt ist. Wenn H der Informationsgehalt in bit/Begriff ist, so ist $Hz_o = K_k$ bit. Aus der Forderung $z \leq z_o$ ergibt sich also

$$\frac{M}{I} = v \leq \frac{K_k}{H} (1 - \varrho)$$

(K_k nach H. Frank, 1960).

Man darf also den Begriffsfortschritt nicht über einen gewissen Maximalwert steigern, wenn die Kapazität des Bewußtseins nicht überschritten werden soll. Je größer ϱ ist, umso kleiner muß v sein. Für I gilt:

$$I \geq \frac{MH}{K_k (1 - \varrho)} = \frac{M}{z_o (1 - \varrho)} .$$

Ein Programm für M Lehrschritte muß also eine Mindestzahl

$$I_o = \frac{MH}{K_k(1 - \varrho)} = \frac{M}{z_o(1 - \varrho)}$$

Frames haben.

4. Der Abbau der semantischen Information der Begriffe

Wenn ein Begriff im Mittel die semantische Nachrichtenmenge H bit enthält, so ist bei der erstmaligen Erwähnung dieser Wert als subjektive Information des Begriffes anzusprechen. Bei jeder Erwähnung des Begriffes wird nun im Mittel der subjektive Informationsgehalt des Begriffes um die (einstweilen unbekannte) Größe k sinken. Der Lernfortschritt (im Sinne der von Cube´schen Redundanztheorie des Lernens (1963)) ist also pro Frame $\Delta R = z.k$. Die Zahl r der Wiederholung muß also mindestens so groß sein, daß $r \Delta R = rzk = H$ ist. Es ist also

$$r = H/k \cdot z ,$$

wenn der Begriff vollkommen gelernt worden ist. Das entspricht der These von Ludwig (1965), daß die Zahl der Repetitionen eines Begriffes ein Maß für seinen semantischen Informationsgehalt ist. Programme mit verschiedenen z dürfen jedoch wegen der Abhängigkeit der Größe r von z nicht mit der gleichen Repetitionszahl r der Begriffe aufgebaut werden. Hier muß zum Vergleich der Schwierigkeit der Begriffe die Größe $rz = H/k$ herangezogen werden, die wie in Abschnitt 2 gezeigt wurde, gleich der Zahl derjenigen Begriffe ist, die mit dem Lehrbegriff z u s a m m e n e r w ä h n t werden, d.h. aber im wesentlichen mit der Zahl der Begriffe, die ihn logisch definieren. In verschiedenen Programmen muß rz eine Konstante - ein Sollwert - sein, der nur von psychologischen Daten der Adressaten und vom mittleren Informationsgehalt H abhängt.

Wenn in einem Programm $rz > \frac{H}{k}$ ist, dann hat der Autor H zu groß eingeschätzt.

Ist $rz < \frac{H}{k}$, so hat er H zu klein eingeschätzt und der Informationsgehalt wird nicht vollständig abgebaut, der Lehreffekt ist schlecht, während bei $rz > H/k$ das Programm langweilig ist, weil der Schüler Tatsachen erneut erfährt, die er schon weiß.

Nun ist bedauerlicherweise H/k nicht bekannt, und es sind auch keine Möglichkeiten zu sehen, wie es absolut bestimmt werden könnte.

Wenn wir von der Hypothese ausgehen, daß "die Autoren" im Mittel gute Programmierer sind, kann man H/k als den Mittelwert von r z über alle Programme, die existieren, ansetzen. Es wird dabei der "mittlere Programmierer" definiert, der H ebenso oft unterschätzt wie überschätzt.

Nun ist $r = z/v$ also $rz = z^2/v$.

Für eine repräsentative Stichprobe von untersuchten Programmen (vergl. Tabelle 2) ist dieser Wert in Tabelle 1 aufgetragen. Der Mittelwert ist $rz = 21{,}4 \pm 20\ \%$. Die relativ geringe Streuung dieses Wertes erscheint in Anbetracht der Verschiedenartigkeit dieser Programme recht bemerkenswert, und man darf daraus ableiten, daß unser "mittlerer Programmierer" eine brauchbare Vorstellung ist.

TABELLE 1

Progr. Nr.	$z^2/v = rz$	$rz - \overline{rz}$	μ
1	25,7	4,3	+ 0,20
2	23,4	2,0	+ 0,09
3	22,4	1,0	+ 0,05
4	16,8	- 5,2	- 0,24
5	21,2	- 0,2	- 0,01
6	16,4	- 5,6	- 0,26
7	28,6	+ 7,2	+ 0,34
8	16,9	- 4,5	- 0,21

$$\sum z^2/v = 171{,}4$$

$$\overline{rz} = 21{,}4 \pm 4{,}35 = 21{,}4 \pm 20\ \%$$

Die Größe $\frac{rz - \overline{rz}}{\overline{rz}} = \mu$ wollen wir die Überschußerwähnung in % nennen.

Es muß erwähnt werden, daß diese Berechnung mit dem "mittleren Programmierer", der seine Adressaten kennt, natürlich nur eine vorläufige Annahme sein kann. Sicher ist z.B. die Konstante k des Adressaten z.B. altersabhängig, so daß $\mu < 0$ auch bedeuten kann, daß das Programm für ältere oder intelligentere Adressaten richtig ist. Immerhin wird man vorläufig beim Programmieren richtig liegen, wenn man $\mu = 0$ anstrebt, also $rz = 21{,}4$.

Nun gibt es unter Programmierern den Begriff des "zerschnittenen Lehrbuchs", mit dem ein schlechtes Programm bezeichnet wird.
Wenn man einen Lehrbuchtext in Abschnitte unterteilt, so kann man diesen unterteilten Text als ein gestrecktes Programm betrachten mit den Abschnitten als Frames. Für diesen Text kann man das (m, i)-Diagramm aufstellen und die Zahlen ϱ und μ ermitteln. Man erhält dann eine Vergleichsmöglichkeit zwischen Lehrbuchtexten und wirklichen Programmen.

Dieses Verfahren birgt natürlich allerlei Probleme. Die wichtigste Schwierigkeit ist die, in welche Stücke der Text zerschnitten werden soll. Wenn man vergleichbare psychologische Lerneinheiten haben will, so müßte man wissen, wieviel Wörter, Sätze oder Abschnitte des Textes von einem Adressaten durchschnittlich gelesen werden, ehe er eine Denkpause macht. Wir nehmen vereinfachend an, daß man grammatische Trennstellen als Framegrenzen heranziehen kann. Eine Denkpause wird ohnehin mit großer Wahrscheinlichkeit mit einem Satzende zusammenfallen. Die einfachste Möglichkeit ist hier

1 Satz = 1 Frame zu rechnen.

Eine andere Möglichkeit ist die, sinnvolle Satzgruppen als Lehrschritte zu zählen. Insbesondere eignen sich für diese Art der Textzerlegung Lehrbuchtexte aus Lehrbüchern, die schon an sich in relativ kurze Sinnabschnitte zerteilt sind. Hier kann man vermuten, daß der Autor sich über eine sinnvolle Absatzgebung bereits Gedanken gemacht hat und den Schüler durch diese Absätze zwingen will, an den betreffenden Stellen Denkpausen zu machen. Ein Lehrbuch, welches derartig kurze Abschnitte hat, ist das Physiklehrbuch "Grimsehl, Bd. 1" (Klett-Verlag). Es wurden für den Abschnitt "Photometrie" dieses Lehrbuchs (m, i)-Diagramme hergestellt, und zwar

1. Nach Sätzen zerlegt,
2. Nach Abschnitten zerlegt.

Die Ergebnisse zeigt Tabelle 2, die mit Tabelle 3 verglichen werden möge.

TABELLE 2

zerlegt nach	v	z	i	ϱ	μ
Sätzen	1,7	3,2	17	0,47	- 0,73
Abschnitten	3,4	5,8	11	0,41	- 0,54

Nr.	Verfasser	Titel	Mittlerer Begriffs-fortschritt Begr./Frame v	Mittlerer Begriffs-inhalt Begr./Frame z	Frame-zahl i	Relative Begriffs-redundanz $= \frac{z-v}{z}$	Überschuß-erwähnung %. 10^{-2} μ	Repeti-tionen r
1	Eher	Elektro-technik. Das Atom	0,56	3,8	104	0,85	+ 0,2	6,8
2	Anschütz	Lichttech-nik	0,72	4,1	150	0,82	+ 0,09	5,8
3	Zielinski	Suezkanal	1,45	5,7	75	0,75	+ 0,05	3,9
4	H. Lindner	Mengen-algebra	0,73	3,5	95	0,79	- 0,24	4,8
5	Zielinski	Meißel-schneide	1,00	4,6	47	0,78	- 0,01	4,6
6	Ludwig	Wetter u. Mensch +)	1,77	5,4	9	0,67	- 0,26	3,0
7	Zielinski	Viskosität	0,77	4,7	47	0,82	+ 0,34	6,1
8	Becher (Westermann)	Binäres Zah-lensystem	4,7	8,9	18	0,47	- 0,21	1,9

TABELLE 3

+) Dieses Programm wurde für auditive Präsentation entworfen.
Die unmittelbare Vergleichbarkeit der Zahlenwerte für seinen eigentlichen Verwendungszweck ist daher nicht gesichert.

Schrifttumsverzeichnis

Frank, H. — Kybernetische Grundlagen der Pädagogik, Agis, Baden-Baden, 1962

Frank, H. — Über grundlegende Sätze der Informationspsychologie, Grundlagenstudien aus Kybernetik und Geisteswissenschaft 1/1, 1960

von Cube, F. — Die Redundanztheorie des Lernens und ihre Anwendung bei Lehrmaschinen. In H. Frank (Hsg.) "Lehrmaschinen in kybernetischer und pädagogischer Sicht", Klett-Oldenbourg, Stuttgart - München 1963

Ludwig, E. — Die Technik zur Herstellung von LP für die programmierte Unterweisung (Henn 1965)

Eingegangen am 14. Januar 1965

BIOLOGISCHE "SOLL"-WERTE ALS STATISTISCHE GRÖSSEN

von Martin Hengst, Berlin

1. Vorbemerkung

Lebende Organismen reproduzieren bestimmte Merkmalswerte trotz wechselnder Umwelteinflüsse mit einer erstaunlichen Präzision. Beispiele hierfür sind die Konstanz von Blutdruck, Körpertemperatur oder Herzfrequenz der einzelnen Individuen. In Analogie zur Technik versuchen wir diese Konstanz biologischer Größen als Ergebnis von Regelungsvorgängen zu verstehen, indem wir uns Organismen als eigengesetzliche Systeme sehr vieler Elemente und Elementarprozesse vorstellen, zwischen denen kreiskausale Zusammenhänge bestehen, die also gewissermaßen rückgekoppelt und dadurch weitgehend von systemfremden Einflüssen unabhängig werden. Dieses als Regelkreislehre bezeichnete Denkmodell (Schmidt 1964) versucht, biologische Systeme und ihre Verhaltensweisen ohne teleologische Begriffe zu beschreiben. Trotzdem hat man in einer etwas kurzschlüssigen Analogie zu technischen Reglern die erwähnten biologischen Größen als "Soll"-Werte bezeichnet und damit überflüssigerweise den Teleologiebegriff wieder ins Spiel gebracht. Es erscheint uns aber gerade ein Verdienst der Regelkreislehre, ihre Denkmodelle konsequent auf dem Begriff der Kausalität aufzubauen.

Auf die logische Unstimmigkeit des "Sollwert"-Begriffes hat bereits Schaefer (1960) hingewiesen, der biologische Sollwerte als empirische Istwerte auffaßt, "die man ihrer Häufigkeit (d. h. ihrer Normalität) wegen zu Sollwerten erklärt hat." Ein solcher statistischer Aspekt biologischer Größen ist nicht neu. Er begegnet uns bereits 1879 bei Galton und Mac Allister, welche die Variation von Reaktionen auf sensorische Stimuli beschrieben.

Später erkannte Kisskalt (1915), daß die S-förmigen Dosis-Wirkungskurven, die den Prozentsatz positiv reagierender Tiere als Funktion der Dosis wiedergeben, nichts anderes als S(%)-Linien, also integrierte Häufigkeitskurven sind. Rautmann (1921) definierte den empirisch gefundenen häufigsten Wert eines Kollektivs physiologischer Merkmale als Norm. De Lind van Wijngaarden (1926) bewies durch umfangreiche Versuche mit Digitalis an Katzen experimentell den statistischen Charakter pharmakologischer Größen. Seitdem haben zahlreiche Forscher durch experimentelle und theoretische Arbeiten die Methoden zur Bestimmung pharmakologischer Normen weiter ausgebaut. Insbesondere vereinfachte Gaddum (1933) die grafische Darstellung der Dosis-Wirkungskurven und damit auch die Auswertung pharmakologischer und serologischer Untersuchungen. Schließlich

wies Hengst (1938) darauf hin, daß jede biologische Norm als statistische Größe aufgefaßt und gemessen werden kann. Er verglich den physiologischen Normalzustand mit einem statistischen Gleichgewicht, bei dem sich der Zustand zwar fortwährend verändert, aber um eine Mittelwertslage pendelt, von der er sich nur selten weit entfernt. Die Kenntnis der Gleichgewichtslage allein genügt noch nicht, um die biologische Norm hinreichend zu kennzeichnen; man muß auch wissen, wie sich die Einzelwerte um die Norm (Gleichgewichtslage) verteilen und die Weite des Pendelausschlages kennen.

Biologische Normen sind also keine Konstanten, sondern variieren von Organismus zu Organismus und - bei ein und demselben Organismus - mit der Zeit. Sie sind gewissermaßen "Eigenwerte" des betreffenden biologischen Systems, die allein mit seiner Stabilität verträglich sind. Ihre Schwankungen sind unvermeidliche Folgen der sie erzeugenden Regelungsmechanismen. Wacholder (1952) führt insbesondere die intraindividuelle Variabilität, d.h. also die Schwankungen der Werte beim einzelnen Lebewesen auf die allgemein-physiologische Organisation aller Lebewesen zurück, die nach ihm auf dem Prinzip eines dynamischen Gleichgewichts und der Rücksteuerung auf dieses Gleichgewicht bei allen Abweichungen beruht. Diese Rücksteuerung erfolgt bei niederen Lebewesen allein schon auf Grund der Eigentümlichkeiten des Stoffwechsels wie z.B. der Reversibilität oder des Massenwirkungsgesetzes, bei höherer Organisation dazu noch über besondere humorale und nervöse Regulationsmechanismen.

2. Verteilungsgesetze biologischer Normen

Die Variationen biologischer Größen führt sofort zur Frage nach einem Verteilungsgesetz, dem sie gehorchen müssen, damit man überhaupt von einer Norm sprechen kann.

Ordnet man Meßzahlen biologischer Merkmale einer bestimmten Art oder Gattung der Größe nach, so wird man sehr oft feststellen, daß an den beiden Enden der Zahlenfolge die extremen Werte relativ selten und in immer größeren Abständen auftreten, während sich die Beobachtungsdaten nach der Mitte zu drängen. Faßt man die Meßzahlen nach Gruppen von mehreren Meßeinheiten zusammen, so sind diese Klassen nach den Enden zu nur vereinzelt, in der Mitte dagegen immer stärker besetzt. Stellt man eine solche Meßreihe, die an etwa 50 oder 100 Organismen gleicher Art gewonnen wurde, in der Weise dar, daß man in der Waagerechten die Meßwertklasse abträgt, in der Senkrechten dagegen die Häufigkeit der Besetzung jeder Klasse, und verbindet man dann die Häufigkeitswerte durch eine Linie, so erhält man sehr oft eingipflige Kurven, die im Idealfall glockenförmig, häufig jedoch auch asymmetrisch verlaufen (Abb. 1).

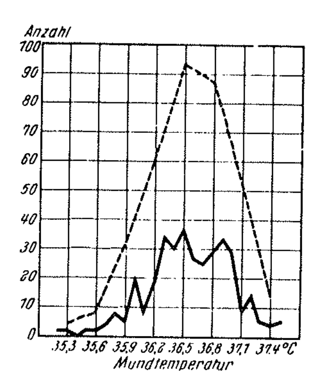


Bild 1

Solche eingipfligen symmetrischen oder asymmetrischen Kurven entstehen erfahrungsgemäß immer dann, wenn für die Ausprägung der kennzeichnenden Eigenschaften tatsächlich nur e i n Haupteinfluß maßgebend war, und alle anderen Einflüsse als Nebenursache nur ein Pendeln um diesen Wert bewirken. Mehr - gipflige Häufigkeitskurven deuten dagegen auf mehrere Haupteinflüsse neben vielen Nebenumständen hin. E i n g i p f l i g k e i t der Häufigkeitskurven ist demnach eine n o t w e n d i g e - wenn auch durchaus nicht h i n r e i c h e n d e - Bedingung für das Vorherrschen e i n e s Hauptgrundes, d. h. für die Existenz einer Norm und damit für die Reproduzierbarkeit unserer Beobachtungen. Eine ausreichende Kennzeichnung der jeweiligen Beobachtungswerte besitzen wir aber zweifellos erst dann, wenn wir ihr Verteilungsgesetz wenigstens angenähert kennen.

Dieses Verteilungsgesetz läßt sich in doppelter Weise darstellen und formulieren: als H ä u f i g k e i t s v e r t e i l u n g und als S u m m e n f u n k t i o n, die nichts anderes als eine aufsummierte Häufigkeitsverteilung ist. Die Häufigkeitsverteilung zeigt uns, welcher relative Anteil der Beobachtungsdaten z w i s c h e n zwei Grenzwerten liegt, die Häufigkeitssummenverteilung den Anteil b i s zu einem Grenzwert. Beide Formen des Verteilungsgesetzes haben ihre Vorzüge und Nachteile: Einheitlichkeit und Symmetrie der Kollektive lassen sich einprägsam durch Häufigkeitskurven veranschaulichen, ihre Ausbreitung wird - besonders bei asymmetrischer Struktur - besser durch Summenkurven dargestellt. Summenkurven gestatten außerdem auch die Darstellung und Auswertung reduzierter Verteilungstafeln mit ungleicher Klasseneinteilung und ermöglichen gegebenenfalls

die graphische Ableitung von Häufigkeitskurven mit äquidistanten Klassen. Zufällige Schwankungen, die besonders in den schwach besetzten Randklassen die Einheitlichkeit der Häufigkeitskurven beeinträchtigen können, wirken sich bei Summenkurven in den mittleren und Endklassen weniger stark aus, da sich die Kurve durch die Aufsummierung gewissermaßen selbst glättet.

Da sich Gestalt, Lage und Ausbreitung als wichtige Eigenschaften von Verteilungskurven erwiesen haben, liegt es nahe, sie durch geeignete mathematische Ausdrücke zahlenmäßig exakt zu erfassen. Die Gestalt läßt sich exakt nur durch den analytischen Ausdruck einer Verteilungsfunktion $y = f(x)$ wiedergeben, Lage und Ausbreitung versucht man durch geeignete Kenngrößen (Parameter) wie Mittelwert und Streuung möglichst eindeutig zu erfassen.

Theoretische Verteilungen (Verteilungsfunktionen) lassen sich unter gewissen Voraussetzungen, also unter der Annahme sogenannter Modelle (Versuchsanordnungen, Urnenziehungen), theoretisch ableiten, wie z.B. die sogenannte Normalverteilung. Nun bewähren sich Gaußsche Normalverteilungen zwar recht gut in der Physik und Technik, versagen jedoch oft bei der Darstellung biologischer und medizinischer Größen, die in den meisten Fällen mehr oder weniger asymmetrische Verteilungen bilden, worauf bereits Fechner (1897) nachdrücklich hingewiesen hat. Es hat sich jedoch gezeigt, daß auch schiefe Häufigkeitsverteilungen symmetrisch werden, wenn statt der gemessenen Werte ihre Logarithmen als Abszissen aufgetragen werden. Wir sprechen dann von logarithmisch-normalen bzw. lognormalen Verteilungen, deren Theorie eingehend von Aitchison und Brown (1957) sowie kurz von Gebelein und Heite (1951) behandelt worden ist.

Wir könnten uns natürlich damit begnügen, die Genauigkeit, mit der irgendeine Häufigkeitskurve empirische Daten beschreibt, als ausreichendes Kriterium ihrer Brauchbarkeit zu betrachten. Es gibt jedoch zwei wichtige Gründe, nach theoretischen Ableitungen von Häufigkeitskurven zu suchen: einmal möchten wir z.B. eine klarere Einsicht in die zugrundeliegenden Prozesse gewinnen. Solche Kenntnisse werden zu weiteren Anwendungen der Verteilungsgesetze anregen. Außerdem erleichtert uns die Kenntnis der elementaren Voraussetzungen, unter denen das betreffende Häufigkeitsgesetz abgeleitet wird, seine eventuelle Modifizierung, um es neuen und anderen Sachverhalten anzupassen. Aus diesen Gründen ist es oft befriedigender, ein Häufigkeitsgesetz zu benutzen, für das wir eine plausible Herleitung kennen, und wir uns nicht nur auf eine mehr oder weniger zufällige empirisch gefundene phänomenologische Beschreibung der empirischen Daten stützen müssen.

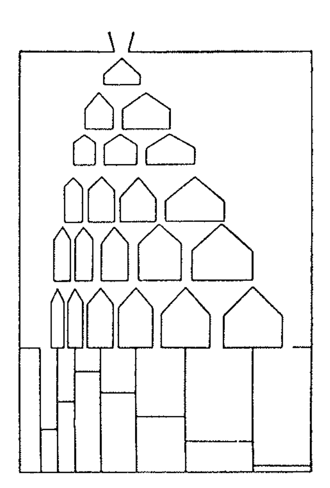


Bild 2

Kapteyn (1903) hat eine Vorrichtung konstruiert (Abb. 2), die - ähnlich dem Galtonschen Brett - die Entstehung schiefer Kurven veranschaulicht. Er nimmt dabei an, daß jede neu hinzutretende Abweichung von der Gleichgewichtslage um so kleiner ist, je stärker das Merkmal bereits vom Mittelwert abweicht. Diese Annahme ist offenbar für rückgesteuerte Systeme sehr plausibel, bei denen sicher die Wahrscheinlichkeit, von der Gleichgewichtslage stark abweichende Zustände anzunehmen, mit zunehmender Abweichung immer geringer wird.

3. Das Wahrscheinlichkeits-Netz (WN)

Um empirische Größen darauf zu prüfen, ob sie einem normalen oder lognormalen Verteilungsgesetz gehorchen, bedient man sich zweckmäßigerweise graphischer Methoden.

Die S-förmige Prozentsummen-Kurve der Normalverteilung läßt sich - besonders im Bereich der kleinen und großen Prozentsummen - nur schwer genau zeichnen. Man benutzt daher nach Henry (1894) für ihre Darstellung Koordinatenpapiere, deren Ordinaten nach dem Gaußschen Integral so geteilt (Abb. 3) sind, daß Prozentsummenkurven von Normalverteilungen zu Geraden gestreckt werden. Anfang und Ende der Prozentsummengeraden lassen sich allerdings nicht zeichnen, da die normale Häufigkeitskurve ja im negativen Unendlichen beginnt und im positiven Unendlichen erst endet. Im allgemeinen beginnen die Ordinaten der Wahrscheinlichkeits-Netze bei 0,02 % und enden bei 99,8 %.

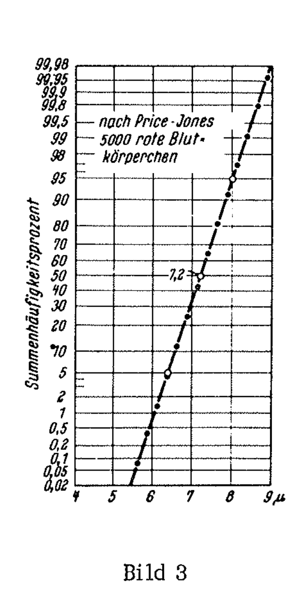


Bild 3

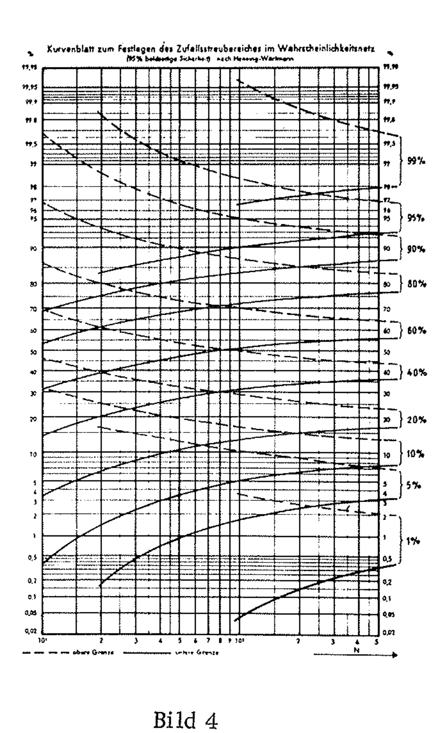


Bild 4

Die Auswertung empirischer Prozentsummenverteilungen im WN bietet eine Reihe von Vorteilen:

1. Man erkennt leicht, ob die Punkte annähernd auf einer Geraden liegen, d. h. ob überhaupt eine Normalverteilung vorliegt und Mittelwertbildung und Streuungsmaß sinnvoll sind.

2. Man benötigt für die Auswertung im WN nicht unbedingt äquidistante Merkmalseinteilungen (gleiche Klassenbreiten).

3. Mittelwert und Streuung lassen sich ohne Rechnung bequem aus der graphischen Darstellung - Zeichnung der "besten" Geraden - im WN ablesen.

Hat man die zu den einzelnen Meßwerten x_i gehörenden Prozentsummenwerte in das WN eingetragen, so legt man - mehr oder weniger willkürlich nach Augenmaß - eine beste Gerade hindurch. Die Abszisse des Schnittpunktes der durch

den 50 %-Ordinatenwert gehenden Parallelen mit der Geraden liefert einen brauchbaren **Schätzwert** für das arithmetische Mittel $\bar{x}$. Entsprechend liest man die Werte $\bar{x}$ - s und $\bar{x}$ + s als Abszissenwerte der Schnittpunkte der 16 % - bzw. 84 %-Linie mit unserer Geraden ab. Die Differenz beider Werte liefert 2s und damit einen Schätzwert für s (vgl. Abb. 3). Wenn also empirisch gewonnene Prozentsummenwerte beim Einzeichnen in das WN bzw. log WN auf einer **Geraden** liegen, so weiß man, daß eine normale Verteilung vorliegt. Je nach dem Umfang der Beobachtungsreihen sind größere oder kleinere Abweichungen von einer Geraden zu erwarten, da ja jede Beobachtungsreihe streng genommen nur eine mehr oder weniger große Stichprobe darstellt. Die für die einzelnen Punkte der Summenprozentgeraden zulässigen Zufallsabweichungen kann man ohne jede Rechnung mit hinreichender Genauigkeit entsprechenden Kurvenblättern entnehmen, die ebenfalls im Handel erhältlich sind (Abb. 4).

Die besondere Bedeutung der Normalverteilung macht es verständlich, daß man asymmetrische Verteilungen durch geeignete Merkmalstransformationen in normale Häufigkeitskurven überzuführen sucht. Das gelingt häufig, wenn man die S(%)-Linie nicht über den oberen Klassengrenzen, sondern über ihren Logarithmen aufträgt. Man benutzt dazu WN, deren Abszissen bereits logarithmisch geteilt sind (vgl. Abb. 5).

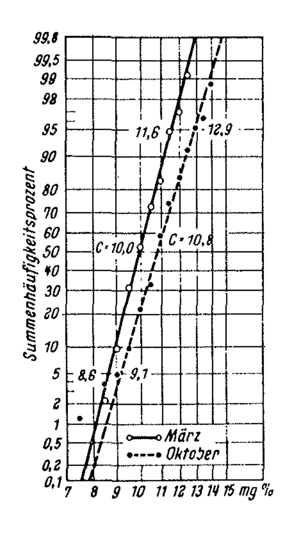


Bild 5

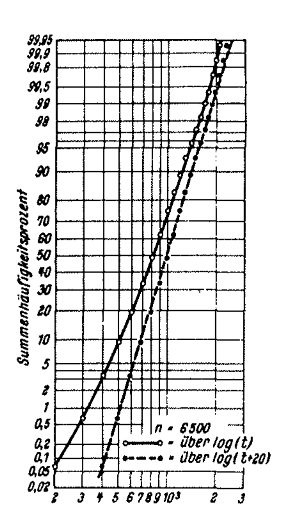


Bild 6

In diesen Fällen erhält man aus den Abszissen der Schnittpunkte unserer Geraden mit der 16 %- bzw. 84 %-Parallelen einen Faktor, durch den man das geometrische Mittel der x_i dividieren bzw. mit dem man es multiplizieren muß, um einen Bereich zu erhalten, in dem wieder 68,3 % aller x_i liegen.

4. Beispiele

Nach den hier geschilderten Methoden sind bereits zahlreiche biologische und medizinische Normen aus empirischen Daten gewonnen worden. So hat z.B. Harting (1952) die Durchmesser von roten Blutkörperchen im WN mit numerischer Merkmalsskala auf Grund von 5000 Messungen nach Price-Jones (1929) dargestellt (Abb. 3). Eine entsprechende Auswertung von Serumkalkwerten im logarithmischen WN teilt Proppe (1952) mit (Abb. 5).

Manche empirische Verteilungen ergeben auch im lognormalen WN noch gekrümmte Kurven. In solchen Fällen erhält man u. U. erst Geraden, wenn man nicht log x, sondern $\log(x \pm c)$ oder $\log(C \pm x)$ auf der Abszisse des WN abträgt. Diese additiven Konstanten c bzw. C geben oft interessante Hinweise über Grenzwerte, die das untersuchte Merkmal nach oben oder unten begrenzen.

So zeigt die Abb. 6 nach Hengst (1952) die Summenprozentlinie der Blutungszeiten nach oberflächlichem Schnitt in das Ohr, die an vollständig normalen Kaninchen von Fleisch und Tripol (1942) bestimmt worden sind, über logarithmischer Scala aufgetragen. Diese Kurve ist im unteren Teil nach links abgebogen. Addiert man vor der Logarithmierung zu den Klassengrenzwerten der angegebenen Blutungszeiten jeweils 20 Sekunden, so erhält man die mit +++++ gezeichnete Gerade. Die Blutungszeiten weisen also dann eine reine Gaußverteilung auf, wenn man als charakteristische Zeit einen um 20 Sekunden vergrößerten Wert wählt. Die Abweichungen der obersten vier oder fünf Punkte von der Geraden deuten auf ein Teilkollektiv hin, dessen Zentralwert doppelt so groß ist wie der des Hauptkollektivs und das evtl. durch eine mechanische Entfernung des Thrombus hervorgerufen sein kann, eine Möglichkeit, auf die bereits die Verfasser hingewiesen haben.

Die Konstante c = 20 läßt sich also graphisch durch ein Näherungsverfahren ermitteln, indem man zunächst versuchsweise einen beliebigen Wert von c wählt.

Erhält man dabei wieder eine nach links oben gekrümmte Linie, so war c noch zu klein; andernfalls erhält man eine nach rechts unten gekrümmte Kurve.

Vielfach beobachtet man auf medizinischem Gebiete asymmetrische Verteilungskurven, die rechtsseitig steil abfallen, die sich also gewissermaßen als Spiegelbilder der bisher behandelten auffassen lassen. Ein anschauliches Beispiel hierzu bildet die Verteilung der Mundtemperaturen von 350 fieberfreien Patienten der Kieler Hautklinik, das Proppe (1952) behandelt. Faßt man die zwischen $35,2^0$ und $37,5^0$ C liegenden Meßwerte in Klassen der Breite $0,3^0$ C zusammen, so erhält man eine Kurve, die über $35,2^0$ und $35,6^0$ bis zu $36,5^0$ flacher ansteigt, als sie von $36,8^0$ auf $37,4^0$ abfällt (Abb. 1).

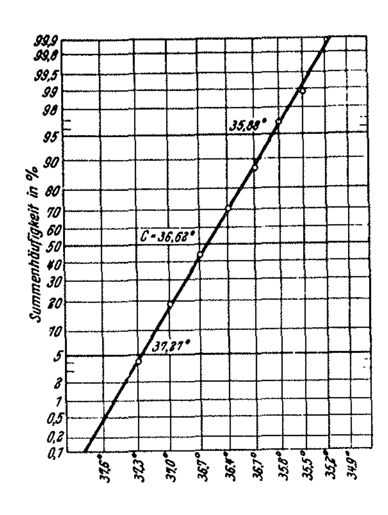


Bild 7

Diese Verteilung läßt sich durch die Transformation log(C-x) normalisieren, wenn wir C etwa gleich 42^0 C setzen (Abb. 7). Die Summenprozentlinie gewinnen wir hier in der umgekehrten Reihenfolge, d.h. wir beginnen mit der Aufsummierung bei den Werten für C - $37,5^0 = 4,5^0$. Der Fluchtpunkt der Verteilung rückt hier also in den Bereich von 41^0 bis 42^0 C. Proppe bemerkt abschließend: "Für die Erkenntnis des Mechanismus der Temperaturregulierung dürfte es dabei bemerkenswert sein, daß die Temperatur des menschlichen Organismus nicht auf einen mittleren Normwert, sondern auf eine Differenz gegenüber einem nicht überschreitbaren Grenzwert einreguliert wird. In ähnlicher Weise läßt sich aus den bei verschiedenen Krankheiten oft sehr unterschiedlichen Häufigkeits-Verteilungen des Sterbealters immer wieder erkennen, daß die Lebensdauer des Menschen sich auf eine Differenz gegenüber dem Grenzwert von etwa 125 bis 130 Jahren einspielt."

Schrifttumsverzeichnis

Aitchison, J. Brown, J.A.	The lognormal Distribution, Cambridge University Press, London 1957
Fechner, G.T.	Kollektivmaßlehre, Leipzig 1897
Fleisch, A. Tripol, J.	Arch. f. exp. Path. u. Pharm. 200, 135 (1942)
Gaddum,J.H.	Med. es. Council 1933, Special Report Series Nr. 183
Gebelein, H. H.-J. Heite	Statistische Urteilsbildung, Berlin 1951
Harting, K. in Daeves, K.	Rationalisierung durch Großzahl-Forschung, Düsseldorf 1952
Hengst, M.	Z. ges. exp. Medizin 104, 89 (1938)
Hengst, M. in Daeves, K.	Rationalisierung durch Großzahl-Forschung, Düsseldorf 1952
Henry, M.P.	Probabilitoe du tir, 1894, Fontainebleau
Kapteyn, J.C.	Skew Frequency Curves in Biology and Statistics, Anstronomical Laboratory, Groningen 1903
Kisskalt, K.	Z.f. Hygiene Bd. 81, 42 (1915)
De Lind van Wijngaarden	Arch. f. exper. Path. u. Pharmklg. 113, 40 (1926)
Price-Jones, C.	J. of. Patholog. a. Bacteriolog. 32, II, 479(1929)
Proppe, A. in Daeves, K.	Rationalisierung durch Großzahl-Forschung Düsseldorf 1952
Rautmann, H.	Untersuchungen über die Norm, Jena 1921
Schaefer, H. in Frank, H.	Kybernetik - Brücke zwischen den Wissenschaften, Frankfurt Main, 1964
Schmidt, H.	VDI-Zeitschrift Bd. 106, 749/53 (1964)
Wacholder, K.	Die Naturwissenschaften 39, 177 (1952)

Eingegangen am 26. Februar 1965

ZUR FRAGE DES AUSWENDIGLERNENS

von Felix von Cube

(Erwiderung auf die Kritik von Herrn Karl Eckel in den GrKG 5/1)

In meiner Arbeit "Über ein Verfahren der mechanischen Didaktik" (von Cube, 1961) habe ich die Theorie aufgestellt, daß sich ein auswendigzulernender Text

$$a_1\ a_2\ a_3\ \ldots\ a_m$$

derart zerlegen läßt, daß seine ursprüngliche Information

$$I_T = m \,.\, \mathrm{ld}m \text{ bit}$$

subjektiv herabgesetzt wird. Zu diesem Zweck wird der Text in q "Wörter" zu je p "Buchstaben" eingeteilt, wobei die Bedingungen

$$q > 1 \quad \text{und} \quad q \,.\, p = m$$

gelten sollen. Die Information eines "Wortes" wird (nach der genannten Theorie) auf diese Weise

$$I_1 = p \,.\, \mathrm{ld}p \text{ bit,}$$

die Gesamtinformation der "Wörter" also

$$I_2 = q \,.\, p \,.\, \mathrm{ld}p \text{ bit.}$$

Für das Zusammenfassen der "Wörter" in der richtigen Reihenfolge wird sodann noch eine zusätzliche Information von

$$I_3 = q \,.\, \mathrm{ld}q \text{ bit}$$

benötigt. Aus der Beziehung

$$I_S = q \,.\, p \,.\, \mathrm{ld}p + q \,.\, \mathrm{ld}q \text{ bit} \qquad (I_S\text{: subjektive Inf.})$$

ergibt sich dann das Minimum von I_S durch die Bedingung

$$\mathrm{ld}p = \mathrm{ld}m - \frac{p-1}{\ln 2} \,.$$

Mit Hilfe dieser Formel kann also ein gegebener Text der Länge m so aufgeteilt werden, daß seine subjektive Information ein Minimum wird.

Herr Eckel machte nunmehr den Einwand, daß die Information eines Textbruchstückes mit p Elementen nicht

$$I_1 = p \cdot \mathrm{ld}p \quad \mathrm{bit}$$

sondern

$$I = p \cdot \mathrm{ld}m \quad \mathrm{bit}$$

betrage; das würde bedeuten, daß jedes der p (zusammengefaßten) Zeichen mit der Information

$$I_Z = \mathrm{ld}m \quad \mathrm{bit},$$

d.h. also mit seiner ursprünglichen Information auftritt.

Der Einwand von Herrn Eckel scheint mir jedoch die Unterscheidung von subjektiver und objektiver Information nicht genügend zu berücksichtigen. Es ist ganz klar (dazu ist die von Herrn Eckel angestellte Rechnung auf S. 32 gar nicht notwendig), daß sich im Bereich der objektiven Information durch keine noch so raffinierte Aufteilung eine Informationsverringerung erzielen läßt: Wenn jedes Zeichen die Information

$$I_Z = \mathrm{ld}m \quad \mathrm{bit}$$

hat, bleibt bei jeder Einteilung die Gesamtinformation des Textes

$$I_T = m \cdot \mathrm{ld}m \quad \mathrm{bit}.$$

Die "mechanische Didaktik" überschreitet indessen den Bereich der objektiven Information; sie will zeigen, daß durch bestimmte Maßnahmen eine gegebene Information subjektiv herabgesetzt werden kann. Wir wollen dies zunächst qualitativ anhand einiger Beispiele erläutern: Hat man ein längeres Gedicht auswendigzulernen, so ist es ein Unterschied, ob man sämtliche Strophen auf einmal ins Auge faßt, oder zunächst nur eine Strophe oder nur eine Zeile. Beim Auswendiglernen einer Zahlenfolge macht es einen erheblichen Unterschied, ob ich für eine längere Folge das Repertoire der zehn Ziffern zugrunde-

legen muß, oder bei einer ersten Teilfolge nur zwei oder drei Ziffern (vgl. hierzu auch die Beispiele von Herrn Eckel, S. 32).

Das Wesentliche bei solchen Vorgängen ist folgendes: Es ist (unter Umständen) von Vorteil, wenn man beim Auswendiglernen eines Textes bewußt auf das gesamte Repertoire verzichtet und sich zunächst nur auf einen noch näher zu bestimmenden Teil beschränkt. Auf diese Weise kann für das den Text lernende Subjekt eine Informationsverringerung erzielt werden.

Es ist Herrn Eckel zu danken, daß er durch seine Kritik auf die (stillschweigende) Voraussetzung, daß das lernende Subjekt diesen Vorgang durchzuführen, d.h. also das Repertoire (wie Herr Eckel sagt) "einzuengen" vermag, aufmerksam gemacht hat. Tatsächlich ist nur unter dieser Voraussetzung (die mir allerdings in der Praxis weitgehend erfüllt zu sein scheint) die subjektive Information eines Textbruchstückes von p Elementen

$$I_S = p \cdot \text{ld}p \text{ bit.}$$

Eine ausführlichere Formulierung der "Theorie der mechanischen Didaktik" müßte demnach folgendermaßen lauten:

Unter der Voraussetzung, daß ein lernendes Subjekt sein (subjektives) Repertoire in einem bestimmten, experimentell noch zu präzisierenden Bereich "einzuengen" (oder aufzubauen) vermag, gibt die "Theorie der mechanischen Didaktik" an, in welcher Weise dies zu geschehen hat, damit eine minimale subjektive Information erzielt wird. (vgl. hierzu auch von Cube, 1964).

Schrifttumsverheichnis

von Cube, Felix — Über ein Verfahren der mechanischen Didaktik, GrKG 2 (1961) H. 1

von Cube, Felix — Kybernetische Grundlagen des Lernens und Lehrens, Stuttgart, 1964

Eckel, Karl — Über den Zusammenhang von "Repertoire" und "Superzeichen", GrKG 5/1 (1964)

Eingegangen am 4. Januar 1965

NACHWORT DER SCHRIFTLEITUNG

Die Schriftleitung möchte mit dieser Erwiderung die Auseinandersetzung um die "mechanische Didaktik" vorläufig abschließen.

Die bisherige Diskussion hat klargestellt, daß sich die durch von Cube vertretene Möglichkeit

$$I_S < m \, \mathrm{ld} \, m$$

nicht schon auf der Ebene der mathematischen Informationstheorie bestätigen läßt, sondern daß sie auf der höheren Ebene der informationellen Systemtheorie diskutiert werden muß, denn es muß wenigstens als Möglichkeit gezeigt werden, wie ein informationsaufnehmendes System S ein geeignet wechselndes Repertoire zugrundelegt.

Offenbar werden für S zunächst zwei Speichereinheiten postuliert: ein "Kurzspeicher", in welchen die Information unmittelbar aufgenommen wird, und ein "vorbewußtes Gedächtnis", in welches sie vom Kurzspeicher her eingelernt wird. Die jeweils erforderlichen Zeiten sind informationsproportional. In diesem Sinne bezieht von Cube die subjektive Information offensichtlich nicht auf die Aufnahmezeit in den Kurzspeicher, denn beim Lesen des ersten Zeichens kann noch nicht bekannt sein, welche p von m verschiedenen möglichen Zeichen im ersten Wort der Länge p vorkommen werden. Die Information pro Zeichen muß also, wie Eckel betont und von Cube nicht bestreitet, für den Kurzspeicher zunächst mit ldm beziffert werden.

Nachdem jedoch die gesamte Zeichenfolge gelesen wurde, ist im Prinzip eine Superierung möglich. Zusätzlich zu den Informationsmengen q . p . ldp und q . ldq muß dann noch festgehalten werden, welches das jeweilige Repertoire von p Buchstaben ist (z.B. durch Speicherung der q verschiedenen Codes). Das Problem, auf welches die Auseinandersetzung führte, lautet also, einen Algorithmus und ein diesen darstellendes System S anzugeben, so daß nicht nur die Einengung auf je eines von q verschiedenen Repertoires zu je p Elementen dargestellt wird, sondern dies auch noch in der Weise geschieht, daß auch unter Mitberücksichtigung der "Umschaltzeit" auf das jeweils neue Repertoire noch die Lernzeit kürzer ist als ohne Superierung. Die Lösung dieses Problems ist sicher nicht in trivialer Weise durch Hinweis auf schon bekannte technische Modelle zu erbringen. Denn für die üblichen technischen Übertragungskanäle bzw. Speicher gibt die "objektive" Information die nicht unterschreitbare Grenze der Übertragungszeit bzw. des Speicherplatzes an, so daß auch durch zusätzliche Übermittlung bzw. Speicherung wechselnder Codes keine weiteren Einsparungen möglich werden.

Die Schriftleitung

UNTERSUCHUNG DER VERMUTUNG J. D. WILSONS ÜBER DEN VERFASSER DES ERSTEN AKTES VON SHAKESPEARES 'KING HENRY VI, FIRST PART' MIT HILFE EINFACHER TEXTCHARAKTERISTIKEN

von Joachim Thiele, Uetersen

John Dover Wilson macht in seiner Edition (1952) der Shakespeareschen Königsdramen auf Grund einer Reihe stilistischer und inhaltlicher Eigentümlichkeiten geltend, daß der ganze erste Akt des ersten Teils von "King Henry VI" und eine Reihe von Szenenfragmenten aus dem dritten und fünften Akt mit hoher Wahrscheinlichkeit von Thomas Nashe [1] geschrieben seien.

Die Bestimmung der beiden Textcharakteristiken $\bar{i}$ und $\bar{r}$ ($\bar{i}$ = mittlere Silbenzahl/Wort, $\bar{r}$ = mittlere Länge der Ketten aus gleichsilbigen Wörtern im Text) einiger Abschnitte von "1 Henry VI", I und einiger Proben aus sicher von Shakespeare stammenden Szenen sowie einiger Abschnitte aus einem kurz nach der Veröffentlichung von "1 Henry VI" geschriebenen Stück Nashes (McKerrow, 1905) ergibt die in Tabelle 1 angegebenen Werte.

Es wird vorausgesetzt, daß zur gleichen Zeit verfaßte Arbeiten eines Autors, die der gleichen Literaturgattung zuzurechnen sind, im Mittel gleiche Textkonstanten aufweisen.

Tabelle 1 und Bild 1 zeigen, daß die Vermutung Wilsons, "1 Henry VI" I. Akt stamme von Nashe, auch aus den statistischen Daten der Texte wohl zu begründen ist.

---

1) Thomas Nashe (oder Nash), geb. 1567 zu Lowestoft in Suffolkshire, Student am St. John's College in Cambridge, lebte seit etwa 1590 in London und starb dort 1600 oder 1601. 1593 schrieb er die satirische Komödie 'Summers last will and testament'; die Probetexte wurden diesem Stück entnommen.

			N[o]	$\bar{i}$	$\bar{r}$
	William Shakespeare				
1	Comedy of Errors I, 1, 2 (1592/93)	I, 1, 31-95	496	1,28	2,54
		2, 1-74	566		
2	1 Henry VI II, 4 (1589/90)	II, 4, 1-61	492	1,25	2,64
3		62-134	572	1,31	2,53
4	3 Henry VI V, 6 (1590/91)	V, 6, 1-43	345	1,28	2,64
		44-93	413		
		Mittelwert		1,28	2,58
	Thomas Nashe				
1	Summers last will (1593)	635-688	424	1,31	2,48
2		689-741	405	1,31	2,28
3		1149-1199	376	1,36	2,26
		Mittelwert		1,32	2,34
	1 Henry VI, I				
1		I, 1, 1-56	431	1,32	2,30
2		108-156	372	1,34	2,09
3		I, 4, 1-56	439	1,33	2,31
		57-111	440		
		Mittelwert		1,33	2,26

[o] N = Anzahl der Wörter des Abschnitts

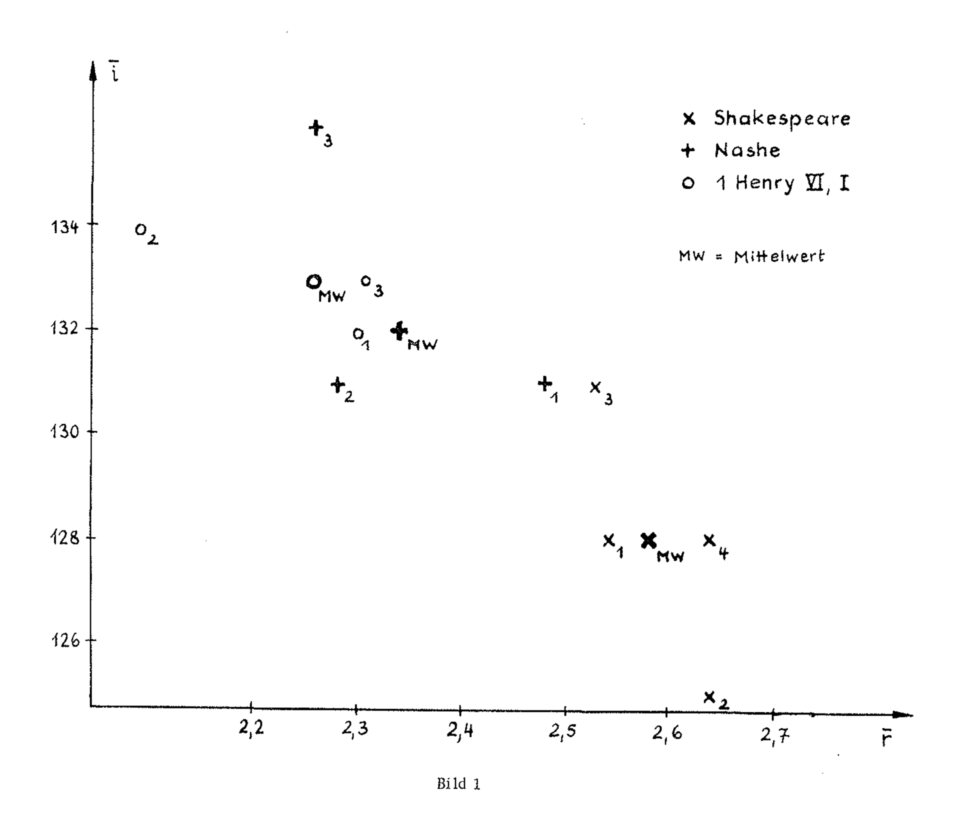


Bild 1

Schrifttumsverzeichnis

Ronald B. McKerrow — 'A Pleasant Comedie, called Summers Last Will and Testament', written by Thomas Nash. London, by Simon Stafford, 1600, abgedruckt in: The Works of Thomas Nashe, edited from the original texts by Ronald B. McKerrow, Text Vol. 3, London 1905, S. 231 ff.

John Dover Wilson — William Shakespeare: The First Part of KING HENRY VI, edited for the syndics of the Cambridge University Press by John Dover Wilson, (Vol. 21 der Gesamtausgabe) Cambridge 1952, S. XXI - XXXI.

Eingegangen am 5. Januar 1965

ZUR WORTRANGKORREKTUR BEI DER AUTOMATISCHEN STICHWORT-ANALYSE

von Helmar Frank (Waiblingen), Berlin

Bergmann, Frank und Großer (1963) waren davon ausgegangen, daß sich der semantische Gehalt eines Textes normalerweise (d.h. wenn nicht der Effekt absichtlich durch geeignete Verwendung von Synonymen und Umschreibungen kompensiert wird) durch eine Verringerung des Ranges derjenigen Wörter syntaktisch äußern werde, die als Stichwörter den Text semantisch kennzeichnen. Diese Annahme hatte sich in erster Näherung bewährt. Es zeigte sich jedoch, daß in sehr vielen Fällen Wortrangverringerungen durch eine zu geringe Textlänge entstehen können. Die Autoren berechneten (übrigens nur näherungsweise, da ihr Rechengang eine nicht genannte Vernachlässigung enthält!) den Erwartungswert der Anzahl derjenigen Wörter, die einen geringeren Rang als ein gegebenes Wort haben würden, aber wegen der Kürze des Textes zufällig nicht auftraten. Aus dem Erwartungswert der so bewirkten Wortrangerniedrigung wurde rückwärts der (im folgenden näher zu definierende) "theoretische Rang" aus dem empirisch bestimmbaren ermittelt. Dieses Verfahren erwies sich zwar als praktisch brauchbar (wenn auch nicht als voll befriedigend), weist aber erhebliche theoretische Mängel auf.

1. Der empirisch bestimmbare Rang eines Wortes W_r erniedrigt sich nicht nur dann, wenn ein an sich wahrscheinlicheres Wort zufällig überhaupt nicht auftaucht, sondern allgemeiner auch dann, wenn es zufällig weniger häufig als W_r benutzt wurde.

2. Andererseits erhöht sich der empirisch bestimmbare Rang eines Wortes W_r, wenn ein an sich unwahrscheinlicheres Wort zufällig häufiger als W_r benutzt würde.

3. Man kann den empirischen Rang r_e eines Wortes gegebenen theoretischen Ranges r_{th} als Zufallsgröße betrachten wie umgekehrt auch den möglichen theoretischen Rang r_{th} eines Wortes vom empirischen Rang r_e. Bestimmt man den jeweiligen Erwartungswert $\bar{r}_e(r_{th})$ bzw. $\bar{r}_{th}(r_e)$, dann erhält man zwei Funktionen, von denen nicht notwendig die eine Umkehrfunktion der anderen ist. Die Autoren benutzten aber die Umkehrfunktion der als $\bar{r}_e(r_{th})$ interpretierten Funktion als $\bar{r}_{th}(r_e)$.

Eine schärfere, aber leider praktisch nicht brauchbare Methode zur Rangkorrektur vermeidet diese Fehler. Sie macht folgende Voraussetzungen:

1. Wenn zwecks Ausdrucks eines Sachverhalts ein Text erzeugt wird, dann wird zunächst willkürlich eine Textlänge N bestimmt.

2. Unabhängig von der gewählten Textlänge wird eine nur vom auszudrückenden Sachverhalt abhängige Urne mit Losen herangezogen, wobei jedes Los genau mit einem Wort W_r beschriftet ist, dasselbe Wort aber auf mehreren Losen stehen kann, derart, daß die relative Häufigkeit $h_r^{(u)}$ des Wortes W_r in dieser Urne sich ergibt aus

$$h_r^{(u)} = k \cdot \left(r + \frac{z}{z-1} \right)^{-\frac{1}{T}}$$

mit

$$\sum_{r=1}^{\hat{r}} h_r^{(u)} = 1 \quad (\hat{r} = \text{Umfang des Wortschatzes}).$$

(z und T sind als feste Konstanten anzunehmen)

3. Der Text wird erzeugt durch Ziehen eines Loses, Abschreiben des darauf stehenden Wortes, Zurücklegen des Loses in die Urne, Mischen, Ziehen des nächsten Loses usf., bis die Textlänge N erreicht ist. W_r wird also für alle r bei jeder Ziehung mit der Wahrscheinlichkeit $p_r = h_r^{(u)}$ gezogen, kann aber bei kleinem N im Text mit einer von p_r beträchtlich abweichenden relativen Häufigkeit $h_r^{(t)}$ auftauchen.

Dieses Modell, das trotz seiner außerordentlichen Grobheit für das Verfahren von Bergmann, Frank und Großer selbst bei einer Anwendung unter zusätzlichen groben Vernachlässigungen brauchbare Resultate lieferte, betrachtet also nicht nur jeden Text als verdampft, sondern setzt voraus, daß der bibliothekarisch relevante Teil seiner semantischen Information bereits in der Verteilung $\{p_r\} \equiv \psi$ stecke. Statt dieser Verteilung kann aber dem empirischen Text nur die Verteilung $\{h_r^{(t)}\} \equiv \varphi$ entnommen werden. Welches ist der Erwartungswert des Ranges r_{th}, den ein Wort in ψ hat, wenn sein Rang in φ als r_e bestimmt wurde? Dieser Zusammenhang stellt die gesuchte Wortrangkorrektur dar. Aus der Untersuchung von Bürmann, Frank und Lorenz (1963) ergibt sich, daß bei wissenschaftlichen Texten z = 2 gesetzt werden darf. Nimmt man die Texttemperatur fest

an (mangels besserer Hypothesen wird man zunächst $T = 1$ setzen), dann kann man k aus $p_1 \approx h_1^{(t)} \approx 0{,}034$ ermitteln, und daraus näherungsweise $\hat{r}$. Da mithin das Verteilungsgesetz in den Urnen für alle Urnen (d.h. entsprechende Sachverhalte) dasselbe ist, unterscheiden sich die Urneninhalte nur noch

a) im Wortschatz (d.h. in der Menge verschiedener Wörter, nicht mehr in der Mächtigkeit dieser Menge!)

b) in der Zuordnung der $\hat{r}$ Ränge zu den $\hat{r}$ Wörtern des jeweiligen Wortschatzes, wobei $\hat{r}!$ Permutationen möglich sind.

Wir setzen nun auch noch den Wortschatz in allen Urnen als übereinstimmend voraus. (Die Transistortechnik und die Bienenzucht kommen also mit demselben Wortschatz aus, nur daß das Wort "Wabe" in der Literatur über Transistortechnik einen etwa gleich extrem hohen Rang hat wie das Wort "Kollektor" in einer Spezialarbeit über Bienenzucht.)

Dann gibt es $\hat{r}!$ a priori gleichwahrscheinliche Verteilungen ψ, d.h. eben diese Zahl möglicher Urnen im mathematischen Modell. Die Wahrscheinlichkeit, daß ein noch völlig unbekannter Text aus der Urne mit der Verteilung ψ erzeugt wurde, ist also

(1) $$p(\psi) = \frac{1}{\hat{r}!} \quad .$$

Die Wahrscheinlichkeit, daß ein Text der Länge N, der aus dieser Urne erzeugt wurde, die empirische Verteilung $\varphi = \{ h_r^{(t)} \}$ aufweist, ist bekanntlich die polynomische Verteilung

(2) $$p_\psi(\varphi) = \frac{N!}{\prod_{r=1}^{\hat{r}} (N \cdot h_r^{(t)})!} \cdot \prod_{r=1}^{\hat{r}} p_r^{N \cdot h_r^{(t)}} \quad .$$

Damit ergibt sich die Wahrscheinlichkeit dafür, daß ein noch unbekannter Text die empirische Verteilung φ hat und aus der Urne mit der Verteilung ψ hervorging, zu

(3) $$p(\varphi\psi) = p(\psi) \cdot p_\psi(\varphi) \quad .$$

Da es nur $\hat{r}!$ Verteilungen ψ gibt, ist hieraus ermittelbar

(4) $$p(\varphi) = \sum_\psi p(\varphi\psi) \, ,$$

womit ferner auch die Wahrscheinlichkeit berechenbar wird, mit welcher der empirischen Verteilung φ die Verteilung ψ in der Urne zugrundeliegt:

(5) $$p_\varphi(\psi) = \frac{p(\varphi,\psi)}{p(\varphi)} .$$

Das Wort, das im empirischen Text den Rang r_e hat, hatte in jeder Verteilung ψ einen von ψ abhängigen "theoretischen" Rang $r_{th}(r_e, \psi)$. Damit ist schließlich der Erwartungswert von $r_{th}(r_e)$

(6) $$\bar{r}_{th}(r_e) = \sum_{\psi} p_\varphi(\psi) \cdot r_{th}(r_e, \psi) .$$

Dies ist die gesuchte Wortrangkorrektur. Da auf der rechten Seite nur Größen stehen, welche durch die Gleichungen (1) bis (5) in jeweils endlich vielen Schritten aus der Statistik des empirischen Textes und aus der kanonischen Verteilung in den Urnen berechenbar sind, ist die Aufgabe theoretisch gelöst. Die explizite Durchführung dieser Methode scheidet natürlich auch bei Verwendung größter Rechenautomaten aus.

Solange nicht abschätzbare (evtl. auch graphische) Näherungen für die Auswertung der Beziehung (6) gefunden werden, müssen wohl für eine praktisch diskutable Lösung des Wortrangkorrekturproblems Vernachlässigungen gemacht werden, die auch bei Anerkennung des angegebenen Modells noch theoretisch bedenklich sind, ohne allerdings so unplausibel sein zu müssen, wie bei der von Bergmann, Frank und Großer benutzten Korrektur. Es mögen hier zwei theoretisch wohl weniger bedenkliche, aber noch nicht praktisch erprobte Alternativen angefügt werden.

1. Man kann (wie bei Bergmann, Frank und Großer) statt mit $\bar{r}_{th}(r_e)$ mit der Umkehrfunktion von $\bar{r}_e(r_{th})$ arbeiten (wobei natürlich jede dieser Funktionen noch von N abhängt). Da die exakte Bestimmung auch dieser Funktion praktisch unüberwindliche Schwierigkeiten zu bereiten scheint, kann (für N als Parameter) der unserem Modell zugrundegelegte Zufallsprozeß auf einem Rechenautomaten simuliert und daraus als statistische Kurve $\bar{r}_e(r_{th})$ gewonnen werden.

2. Man kann nach einem Autor suchen, der mehrfach denselben Sachverhalt mit abweichenden Formulierungen in verschiedenen Publikationen verschiedener Länge dargestellt hat. Die Häufigkeitsverteilung im Mischtext, der aus allen diesen Texten gebildet wird, setzt man dann der als gemeinsam angenommenen Wahrscheinlichkeitsverteilung ψ gleich. Damit ist für jeden Einzeltext bezüglich

des Mischtextes eine eineindeutige Zuordnung zwischen den Rängen r_e und r_{th} der im Einzeltext vorkommenden Wörter gegeben. Man versucht nun eine gewichtete Mittelwertbildung der Form

$$\bar{r}_{th}(r_e) = \sum_{x=-S}^{+S} g(r_e + |x|) \cdot r_{th}(r_e + x)$$

mit irgendeiner Schranke S und einer für wachsendes $|x|$ abnehmenden Gewichtsfunktion $g(r_e + |x|)$ vorzunehmen, so daß eine hinreichend geglättete Kurve durch die empirisch gefundene Punktmenge $\{ r_e \mid r_{th} \}$ gelegt werden kann. Macht man dies für alle verschieden langen Teiltexte, dann entsteht eine Kurvenschar mit dem Parameter N. Dieser Kurvenschar wird man künftig auch für andere Texte vergleichbarer Länge die Wortrangkorrektur entnehmen.

Schrifttumsverzeichnis

Bergmann, Frank und Großer — Die Wortrang-Differenz als semantisches Indiz Grundlagenstudien aus Kybernetik und Geiseswissenschaft 4, Heft 3/4, 1963, S. 91-107

Bürmann, Frank und Lorenz — Informationstheoretische Untersuchungen über Rang und Länge deutscher Wörter Grundlagenstudien aus Kybernetik und Geisteswissenschaft 4, Heft 3/4, 1963, S. 73-90

Eingegangen am 10. Februar 1965

Richtlinien für die Manuskriptabfassung.

Es wird zur Beschleunigung der Publikation gebeten, Beiträge an die Schriftleitung in doppelter Ausfertigung einzureichen. Etwaige Tuschzeichnungen oder Photos brauchen nur einfach eingereicht zu werden.
Artikel von mehr als 12 Druckseiten Umfang können in der Regel nicht angenommen werden. Unverlangte Manuskripte können nur zurückgesandt werden, wenn Rückporto beiliegt. Es wird gebeten bei nicht in deutscher Sprache verfaßten Manuskripten eine deutsche Zusammenfassung anzufügen und wenn möglich, zur Vermeidung von Druckfehlern, das Manuskript in Proportionalschrift mit Randausgleich als fertige Photodruckvorlage einzusenden.
Die verwendete Literatur ist, nach Autorennamen alphabetisch (verschiedene Werke desselben Autors chronologisch) geordnet, in einem Schrifttumsverzeichnis am Schluß des Beitrags zusammenzustellen. Die Vornamen der Autoren sind mindestens abgekürzt zu nennen. Bei selbständigen Veröffentlichungen sind Titel, Erscheinungsort und -jahr, womöglich auch Verlag, anzugeben. Zeitschriftenbeiträge werden vermerkt durch Name der Zeitschrift, Band, Seite (z. B. S. 317–324) und Jahr, in dieser Reihenfolge. (Titel der Arbeit kann angeführt werden). Im selben Jahr erschienene Arbeiten desselben Autors werden durch den Zusatz „a", „b" etc. ausgezeichnet. Im Text soll grundsätzlich durch Nennung des Autorennamens und des Erscheinungsjahrs des zitierten Werkes (evtl. mit dem Zusatz „a" etc.), in der Regel aber nicht durch Anführung des ganzen Buchtitels zitiert werden. Wo es sinnvoll ist, sollte bei selbständigen Veröffentlichungen und längeren Zeitschriftenartikeln auch Seitenzahl oder Paragraph genannt werden. Anmerkungen sind zu vermeiden.

Forme des manuscrits.

Pour accélérer la publication les auteurs sont priés, de bien vouloir envoyer les manuscrits en deux exemplaires. Des figures (à l'encre de chine) et des photos, un exemplaire suffit.
En général les manuscrits qui fourniraient plus de 12 pages imprimées ne peuvent être acceptés. Les manuscrits non demandés ne peuvent être rendus que si les frais de retour sont joints. Si les manuscrits ne sont pas écrits en allemand, les auteurs sont priés de bien vouloir ajouter un résumé en allemand et si possible, pour éviter des fautes d'impression, de fournir le manuscript comme original de l'impression phototechnique, c'est-à-dire tapé avec une machine aux caractères standard et avec marges étroites.
La littérature utilisée doit être citée à la fin de l'article par ordre alphabétique; plusieurs oeuvres du même auteur peuvent être enumérées par ordre chronologique. Le prénom de chaque auteur doit être ajouté, au moins en abrégé. Indiquez le titre, le lieu et l'année de publication, et, si possible, l'éditeur des livres, ou, en cas d'articles de revue, le nom de la revue, le tome, les pages (p.ex. p. 317–324) et l'année, suivant cet ordre; le titre des travaux parus dans de revues peut être mentionné. Les travaux d'un auteur parus la même année sont distingués par «a», «b» etc. Dans le texte on cite le nom de l'auteur, suivi de l'année de l'édition (éventuellement complété par «a» etc.), mais non pas, en général, le titre de l'ouvrage; si c'est utile on peut ajouter la page ou le paragraphe. Evitez les remarques en bas de pages.

Form of Manuscript.

To speed up publication please send two copies of your paper. From photographs and figures (in indian ink) only one copy is required.
Papers which would cover more than 12 printed pages can normally not be accepted. Manuscripts which have not been asked for by the editor, are only returned if postage is enclosed.
If manuscripts are not written in German, a German summary is requested. If possible these manuscripts should be written as original for phototechnical printing, i. e. typed with proportional types and with straight-line margin.
Papers cited should appear in the Bibliography at the end of the paper in alphabetical order by author, several papers of the same author in chronological order. Give at least the initials of the authors. For books give also the title, the place and year of publication, and, if possible, the publishers. For papers published in periodicals give at least the title of the periodical in the standard international abbreviation, the volume, the pages (e.g. p. 317–324) and the year of publication. (It is useful to add the title of the publication.) When more than one paper of the same author and the same year of publication is cited, the papers are distinguished by a small letter following the year, such as "a", "b" etc. References should be cited in the text by the author's name and the year of publication (if necessary followed by "a" etc.), but generally not with the full title of the paper. It might be useful to mark also the page or paragraphe referred to.